Virtualization and reliability

http://fermigrid.fnal.gov

Steven C. Timm
FermiGrid Services Group, Fermilab
with significant contributions from
Dan Yocum and Keith Chadwick

OSG Site Admins @TTU, Aug 2011

Introduction

- FermiGrid has run Highly Available VOMS, GUMS, MySQL, SAZ, and Squid servers for 3 years now, with uptime stretches of 6 months with 99.9952% uptime.
- GUMS and SAZ rates are regularly 18k/hr.
- Available HA Technologies
- Overview of how we made GUMS and VOMS highly Available
- New directions in High Availability

Planning a High Availability System

- Highly available facility, UPS and generator preferable
- TWO highly available facilities, even better.
- Redundant power, two different power panels.
- Heartbeat—ways for a system to check that the other system is alive.
- How to get the ip traffic forwarded to the right system or systems
- How and where is the state of the application stored
- How to keep the multiple masters in sync
- How to maximize throughput and minimize power

Example: GUMS

- All state information stored in MySQL database
- Web service on a single port, easily redirected if necessary
- High availability and high throughput essential
- Develop a shared-nothing system
- (no NFS, no SAN)
- GUMS can run in limited memory, but is one of several tomcat apps that wants the port 8443—good candidate for virtualization
- As a site administrator you should be running GUMS even if you are a small site. We are making the grid-mapfile instructions harder to find every time.

IP HA technologies

- FermiGrid considered three options
  - Round robin DNS
  - Linux Virtual Server
  - Switch-based load balancing.
- Selected Linux Virtual Server
  - Piranha product from Red Hat
  - Use Direct Routing (LVS-DR) method
  - Trying to extend this to IP tunneling.
- Listens on voms.fnal.gov, gums.fnal.gov, saz.fnal.gov, fg-mysql.fnal.gov, squid.fnal.gov, voms.opensciencegrid.org, osg-ress-1.fnal.gov, osg-ress-4.fnal.gov
- Re-directs connections based on IP+port to backend “real servers”
  - Uses weighted least connections (WLC) to schedule connection requests
  - Other algorithms available

HeartBeat (Pulse)

- In LVS-DR, real servers respond directly to client
  - LVS “pings” services for availability
    - Removes service from scheduling if unavailable
    - Adds service back in when available
  - LVS master server is fg5x0
  - LVS backup server is fg6x0
    - Active-passive configuration
    - Failover in 6 seconds “pulse” daemon
    - Active connections to backend servers are maintained
  - If a service or real server fails during an open connection, the connection is lost.

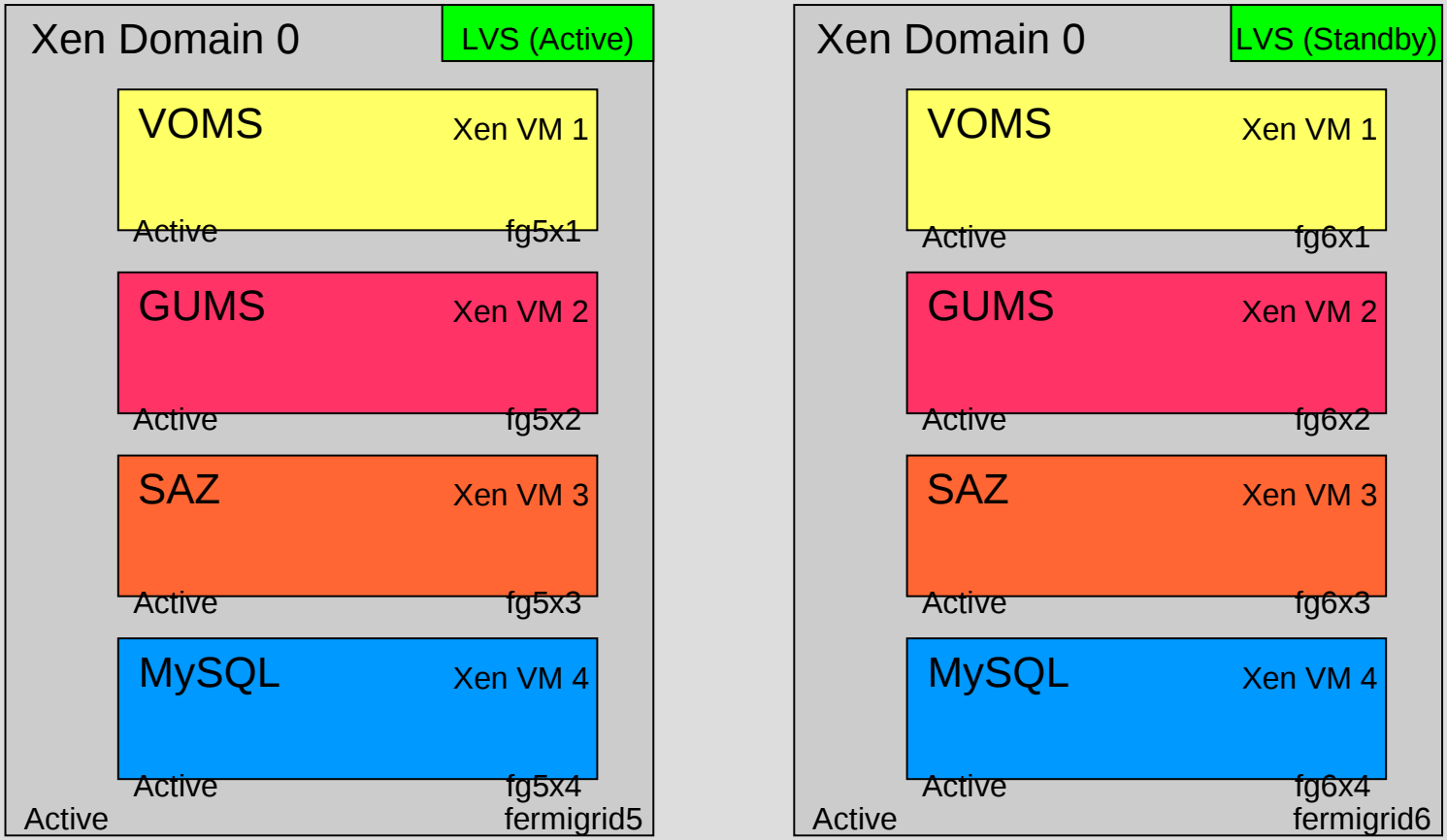
MySQL

- Investigated several technologies
  - MySQL Cluster
  - drbd – distributed remote block device
  - mysqlhotcopy
  - mysqldump + rsync
  - Multi-master replication (requires MySQL >v5.0.2)
- Decided on Multi-master replication
  - Scales well to 10 systems
    - currently only using 2
  - Near real-time replication, 1.1ms for 1KB record
  - Database server outages handled correctly
    - Logs replayed after server rejoins chain

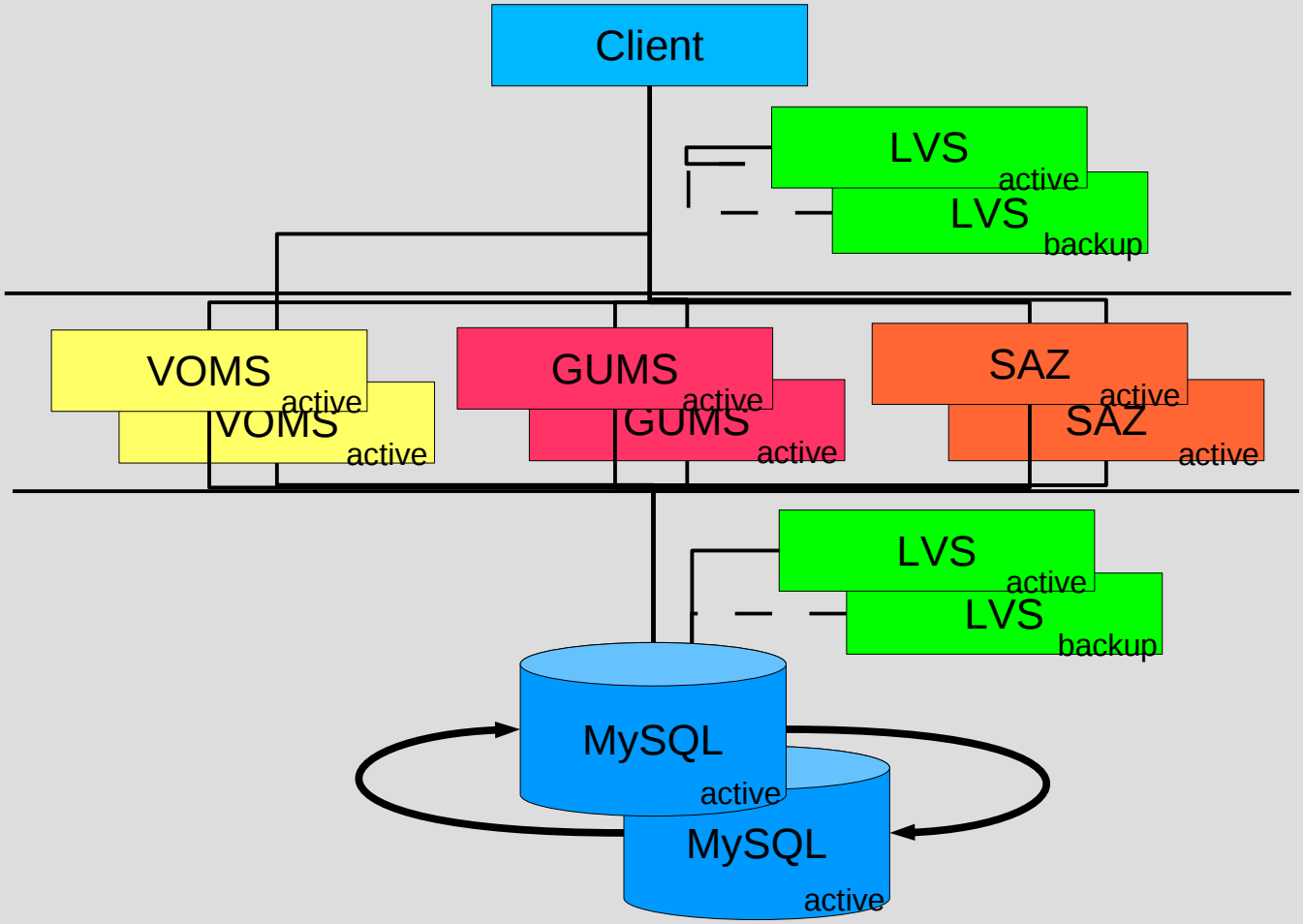
MySQL Replication

- Complete failure requires full copy (obviously)
  - >2 systems require special techniques to close the circular chain in the event of failure
- Service Replication
  - 2 VOMS “servers”, one each on fg5x1 & fg6x1
    - Each machine serves 12 VOs
  - 2 GUMS servers, one each on fg5x2 & fg6x2
  - 2 SAZ servers, one each on fg5x3 & fg6x3
  - All services use replicated MySQL servers on fg5x4 and fg6x4
  - No source code modifications

Organization of VM's



FermiGrid-HA



Note 1: All network connections are on the public network

Note 2: LVS directors displayed separately for convenience – they are the same in reality

Near-trivial HA

- Squid—just set up two of them and front with LVS. No shared FS required.
- Web-services listener for the BDII. No persistent state
  - GOC running this with round robin DNS since 2009.
- ReSS Information Gatherer—No persistent state
  - Osg-ress-1 and osg-ress-4 since 2009.
- Ganglia—just set up two and have gmond send to both
- Condor Collector/Negotiator
  - Condor has its own HAD daemons that can transparently fail over the negotiator between several collector-running machines. All our production clusters running these.
- RSV—just make two of them and set the jobs to offset at different times.

DRBD + Heartbeat

- MyProxy and globus gatekeepers store a lot of state in flat files on disk
- Use DRBD—Dynamic Remote Block Device
- Active-passive configuration. Master writes to its local disk and then a copy of every block is sent across net to remote mirror
- Heartbeat detects if master is down, if so it grabs the service IP and asserts itself as the DRBD master.
- DRBD is kernel-module based solution, available in ATRPMS and SL contrib repos.
- Have been running MyProxy like this for 1+ years
- Making plans to do same for gatekeepers
- New config we haven't tested yet—dual master, GFS on top of DRBD so both sides can write.

FermiGrid HA-2

- From Feb 2010-May 2011, 4 power cuts and 7 network cuts in our “High Availability” building
- We moved half of our services machines to a different building
- Dark fiber to connect private net
- Same LAN subnet for both (may use OTV eventually).
- Move complete on June 7. The same day high temps caused the AC to fail in the new building.
- We can now stay up if there is power or network outage in either building, and we have seen both.
- Now have to get rest of services (Gatekeepers and worker nodes) to take advantage of this.
- Currently heavily dependent on Bluearc NAS, which is only in one building.

Now comes the hard part

- Worker nodes need CA certs-
  - Transfer to fetch_crl v3 with local cached copy of CA certs.
- Worker nodes need non-shared worker node client + gLexec
  - Eagerly awaiting new RPM-based release of WN client.
- Gatekeepers currently depend on NFS home directories
  - Can switch to NFS-lite
- Active-passive failover of gatekeeper requires shared FS
  - DRBD might work, might not
  - Lots of state info, gratia, $GLOBUS_LOCATION/tmp, more.
- NFS-Lite configuration means we have to bring home directories along with us too. (only works for Condor, we also have PBS)
- condor_schedd—has native HA availability but requires shared FS between two machines.
- We have converted condor_schedd and globus gatekeeper to run on “service ip’s” so the service can move from machine to machine.
- Proxies moving across NFS in the clear have been sore point from beginning of OSG with no clear resolution.

Hard Part, continued

- Gratia databases (>1TB) not amenable to multi-master MySQL replication
  - MySQL replication doesn't deal with triggers well.
  - We have one-way replication going, Gratia collectors talk to master, Gratia reporters and backups talk to slave.
  - Gratia reporters, mysql ip's can move via heartbeat.
  - .
- WS-GRAM—Crazy file locking precludes NFS, but all state is in files somewhere.
- Spread out the HA across the Fermilab site in case of a building outage.

Virtualization and HA

- Many OSG services are now well-enough behaved to run in 2GB of RAM or less. (VOMS, GUMS). They are not big cpu loads either, especially not VOMS.
- But they all want the same ports and you want to start them independently
- Current servers have 8 cores, 16GB RAM
- We divide each into 4 or 5 Xen domains

Virtualization and HA continued

- Stock Xen 3.1.2 comes with RedHat/Centos/SL 5 update 2.
- Xen guest reboots in 5-10 seconds
- Base Dell PowerEdge server takes 3-5 minutes.
- We started with four services machines (one for each of voms, gums, saz, mysql) and condensed them into 2 physical servers, with 2 instances of each.
- This summer we used the HA features to move the whole system from old to new hardware without scheduling a downtime.
- In production we have used exclusively “Paravirtualized” Xen hosts.
- Sci. Linux 6 doesn't support Xen hypervisors.. KVM is the present and future.
- KVM good for most stuff, not ready for MySQL and postgres yet.
- SL6 KVM significantly better disk I/O than SL5.

Poor man’s HA

- Split the cluster in half, each half with a head node so you will never lose more than ½ cluster at once with a head node failure.
- Have 2-4 gatekeepers going into a single cluster so that cluster is still accessible. Also a must for throughput to big condor clusters.
- Install your single head node using Xen, it is much easier to move it to another piece of hardware later.

To The Cloud!

- In past year we have deployed FermiCloud as a platform for development, integration and testing.
- Infrastructure-as-a-service means that your virtual machine could be running on any number of different physical hosts.
- We are now also serving small production servers.
- There are several open-source ways to have a working small private cloud. We are using OpenNebula.

Live Migration

- Both Xen and KVM hypervisors support live migration
- With cloud software, or with a proprietary virtualization manager like Oracle VM or XenServer it is routine to use.
- Need shared file system between all systems that host virtual machines—NFS in a pinch but SAN preferred
- Admin-initiated live migration
  - System gets signal to migrate
  - Memory and process stack copied to another VM host
  - Once copy is done, old VM is halted and new one comes up. Network connections stay live, uptime is uninterrupted.
- Next step—use heartbeat to detect failure and bring up machine on other host.

The HandsOn part

- Go to Google Docs
- (request an invite from StevenCTimm@gmail.com if necessary).
- Or grab the exported PDF file that is on the Indico page and/or the Twiki.
- http://docs.google.com/Doc?docid=0AVY11NMVz7mPZGdmNjg0bWZfMmZiNWttbmQ0&hl=en
- Major parts:
  - LVS Server
  - Pulse/heartbeat configuration
  - Multi-master mysql
  - Redundant GUMS.
- I can't finish the whole howto in an hour, don't expect you to.
- But will focus on making and demoing the multi-master mysql.
- Easy enough to make and add the 2nd gums server later.
- If you finish early, you can work on other tutorials.